iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69606

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Voice Based Hybrid Sentiment Analysis on Movie Reviews

Mrs. G. Venkateswari[1], Y. Srinija[2], P. Sasikala[3], T. Hima Deepika[4], P. Sirisha[5]
[1]*MTech (Ph.D.), Assistant Professor, Dept of CSE, Bapatla Women's Engineering College, Bapatla, AP, INDIA*
[2, 3, 4]*BTech, Computer science & Engineering, Bapatla Women's Engineering College, Bapatla, AP, INDIA*

***Abstract:** In the digital era, online movie reviews have become a key platform for audiences to share their opinions and sentiments. While many sentiment analysis systems focus exclusively on text-based data, they often miss the subtle emotional signals conveyed through speech. This paper introduces a voice-based hybrid sentiment analysis model that combines both acoustic features and textual content to enhance sentiment classification accuracy. The system integrates machine learning algorithms such as Support Vector Machine (SVM), Naïve Bayes, and Linear Regression to create a robust hybrid model. Acoustic data is analyzed to extract prosodic and spectral features like pitch, energy, and Mel Frequency Cepstral Coefficients (MFCCs), while Natural Language Processing (NLP) techniques are employed to process transcribed text. By merging both audio and text features, the model improves sentiment polarity detection accuracy. Experimental results on publicly available datasets show that this hybrid approach outperforms traditional single-modality methods. This research emphasizes the value of multi-modal sentiment analysis and paves the way for more emotionally intelligent human-computer interactions.*

I. INTRODUCTION

Conventional sentiment analysis systems primarily focus on text-based data, using Natural Language Processing (NLP) techniques to classify sentiment into categories such as positive, negative, or neutral. However, these methods often overlook the emotional nuances conveyed through vocal characteristics such as tone, pitch, rhythm, and intonation. These paralinguistic features provide valuable insights into the speaker's emotional state, which can significantly improve the accuracy and depth of sentiment interpretation.

This approach addresses the gap between textual and speech-based sentiment analysis by proposing a hybrid model that incorporates both textual and acoustic features from spoken movie reviews.

The system combines machine learning algorithms such as Support Vector Machine (SVM), Naïve Bayes, and Linear Regression to improve sentiment classification accuracy. Acoustic signals are analyzed to extract prosodic and spectral features like Mel Frequency Cepstral Coefficients (MFCCs), pitch, and energy, while the spoken content is transcribed and processed using Natural Language Processing (NLP) techniques.

II. LITERARURE SURVEY

1) B. Manjula et al., "An Overview of Sentiment Analysis: Techniques, Approaches, and Challenges," 2019.
2) M. Govindarajan et al., "Hybrid Approach for Sentiment Classification of Movie Reviews," 2013.
3) A. K. M. Masum et al., "A Comparative Study of Naive Bayes and SVM Algorithms for Sentiment Analysis Using Review Datasets," 2020.

III. EXISTING SYSTEM

The Lexicon-Based Approach relies on predefined sentiment dictionaries, where words are assigned sentiment scores. These scores are then aggregated to calculate the overall sentiment of the text. On the other hand, the Machine Learning-Based Approach uses labeled datasets to train models, enabling sentiment analysis by processing text input and generating outputs categorized as positive, negative, or neutral.

IV. PROPOSED SYSTEM

Voice-based sentiment analysis evaluates the tone and emotion behind the speaker's voice. Using machine learning voice features like pitch, and speech rate can be extracted to assess the emotional state of the speaker.

Voice-based hybrid sentiment analysis on movie reviews refers to the speech processing to determine the sentiment (positive, negative, or neutral) behind a spoken review.
This involves analyzing voice data along with text data to better understand the overall sentiment expressed in the review.

V. IMPLEMENTATION

1) Voice Input: Users provide their movie reviews through spoken input.
2) Speech-to-Text Conversion: Utilize an Automatic Speech Recognition (ASR) tool (e.g., Google Speech-to-Text, Whisper, or (Vosk) to transcribe the spoken reviews into text.
3) Sentiment Analysis: Textual Analysis: Apply NLP techniques using models like BERT, RoBERTa, or LSTM to interpret and classify the sentiment of the transcribed text. Audio Analysis (Optional): Extract vocal characteristics such as pitch, tone, and intensity using tools like OpenSMILE or deep learning methods to infer emotional states.
4) Hybrid Sentiment Detection: Integrate both the textual sentiment and audio-based emotional indicators to generate a final sentiment label—such as positive, negative, neutral, or specific emotions.

VI. MODULES

1) Voice Input Module: Captures user-generated movie reviews through either direct microphone input or uploaded audio recordings.
2) Speech-to-Text Module: Transforms the captured audio into textual format using speech recognition tools such as Whisper or Google's Speech-to-Text API.
3) Text Preprocessing Module: Prepares the transcribed text by removing noise, performing tokenization, and formatting it for sentiment evaluation.
4) Text Sentiment Analysis Module: Applies machine learning or deep learning models like BERT or LSTM to assess the sentiment expressed in the cleaned text.
5) Audio Feature Extraction Module: Analyzes the audio input to extract vocal features such as MFCC, pitch, and energy that are indicative of emotional expression.
6) Audio Emotion Recognition Module: Utilizes the extracted acoustic features to classify the speaker's emotional tone, identifying states such as joy, sadness, or anger.
7) Fusion Module: Combines the sentiment derived from textual analysis and emotional cues from audio to produce a final, unified sentiment outcome.
8) Output Visualization Module: Presents the final sentiment result to the user through a user-friendly display or interface, enabling interpretation and further use.

VII. ALGORITHMS

SVM (Support Vector Machine) is a supervised learning algorithm that aims to find the optimal hyperplane for separating data points from different classes while maximizing the margin between them. Naïve Bayes is a probabilistic classifier that applies Bayes' theorem, assuming that the features are independent of each other, making it particularly well-suited for tasks like text classification. Linear Regression is a statistical technique used to model the relationship between input variables and a continuous target by fitting a linear equation to the observed data

VIII. TECHNIQUE

Scikit-learn is a widely used Python library that offers efficient tools for machine learning tasks such as data preprocessing, modeling, and analysis. TensorFlow, developed by Google, is an open-source framework designed for building and training advanced neural networks and machine learning models. The Natural Language Toolkit (NLTK) is a comprehensive Python package for processing human language, providing functionalities like tokenization, stemming, tagging, parsing, and semantic analysis.
Google Speech-to-Text is a cloud-based service that utilizes advanced speech recognition techniques to convert spoken input into written text with high accuracy.

IX. SYSTEM ARCHITECTURE

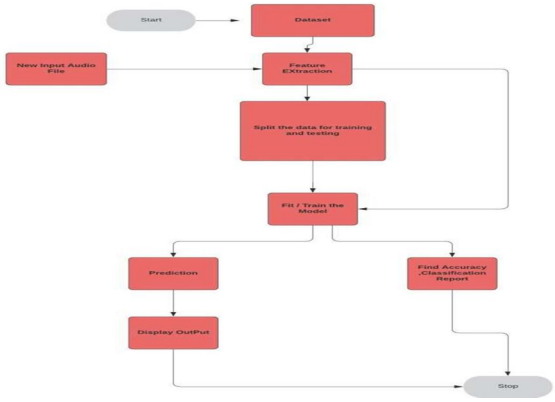


Fig 1

X. DATAFLOW DIAGRAM

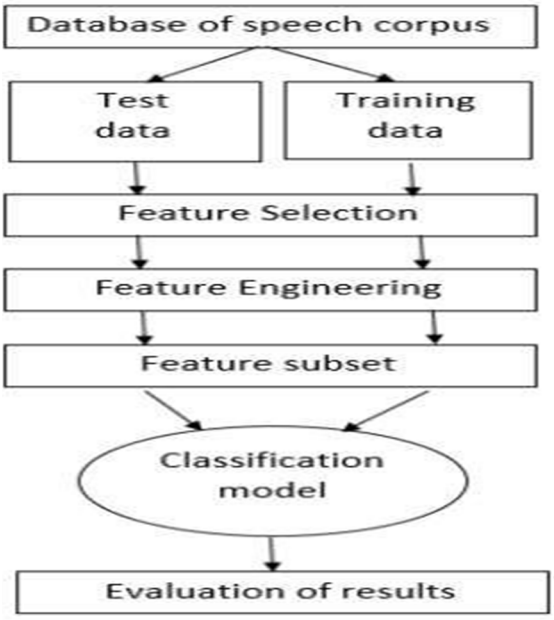


Fig 2

XI. RESULT AND ANALYSIS

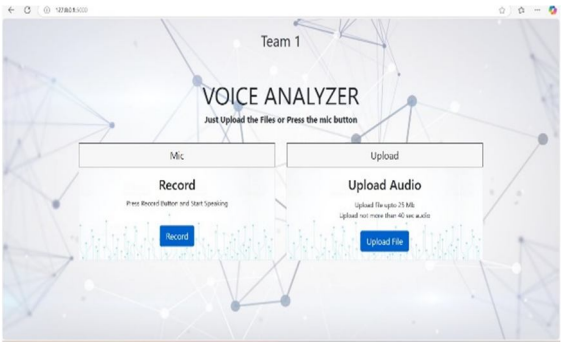


Fig 3

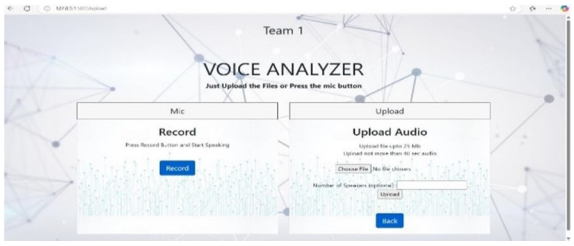


Fig 4

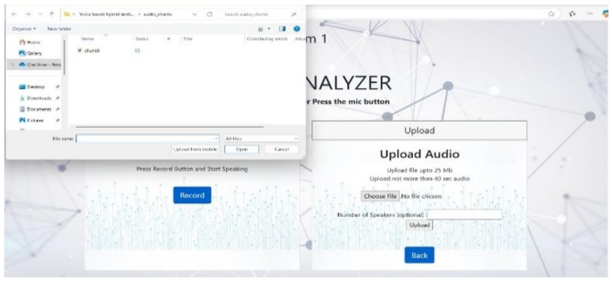


Fig 5

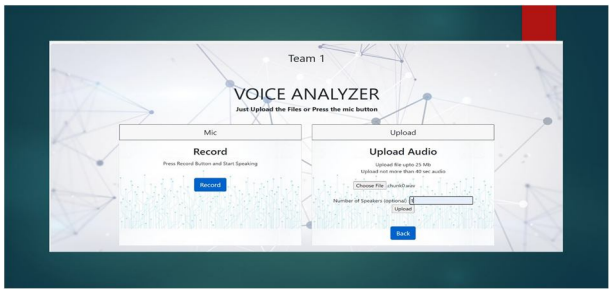


Fig 6

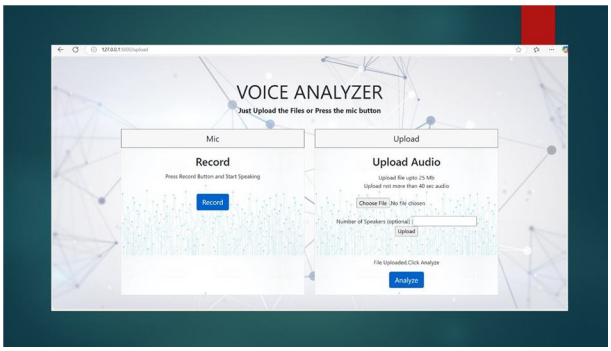


Fig 7

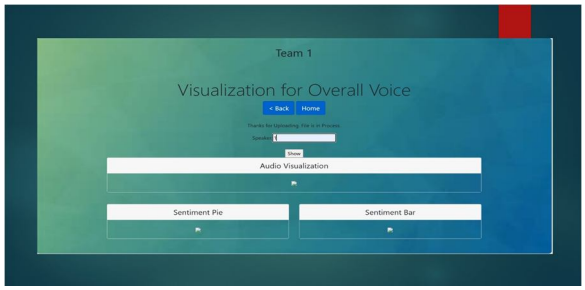


Fig 8

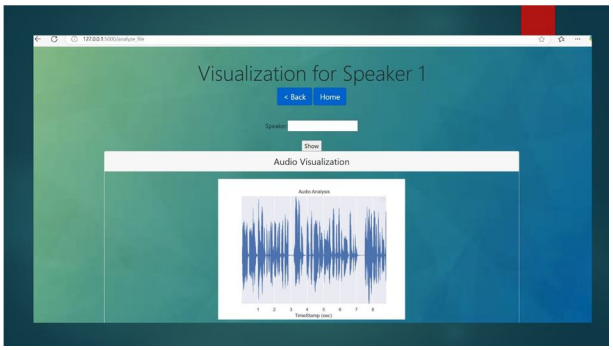


Fig 9

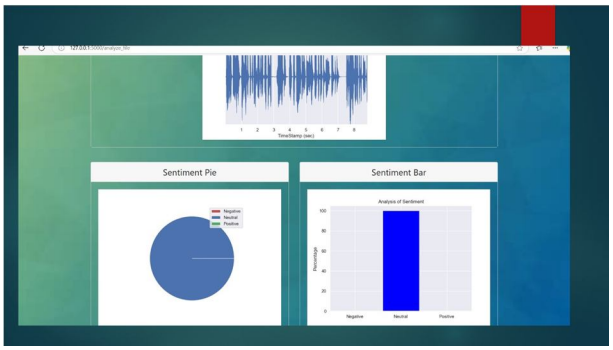


Fig 10

XII. CONCLUSION

The development of a voice-based hybrid sentiment analysis system for movie reviews demonstrates the potential of combining acoustic and linguistic features to achieve more accurate and insightful sentiment detection. By leveraging both the emotional tone in speech and the textual content obtained through automatic speech recognition, the system offers a richer understanding of user opinions. This approach not only enhances sentiment classification performance but also opens new avenues for applications in voice-driven interfaces, entertainment feedback systems, and personalized user experiences. As technology evolves, integrating advanced models and real-time capabilities will further strengthen the effectiveness and adaptability of such systems**.**

XIII. FUTURE SCOPE

The future of voice-based hybrid sentiment analysis for movie reviews offers significant opportunities for growth and innovation. Potential advancements include the incorporation of deep learning algorithms to enhance detection accuracy, multilingual support to reach a wider audience, and real-time processing capabilities for instant sentiment recognition. Additionally, the system could be refined to identify specific emotional states, operate effectively in noisy environments, and deliver personalized sentiment insights. Such developments would greatly expand its usefulness in applications like virtual assistants, content recommendation systems, and real-time audience feedback tools in both entertainment and marketing industries**.**

REFERENCE PAPERS

[1] Wöllmer, M., Eyben, F., Schuller, B., & Rigoll, G. (2013). Sentiment Analysis in Audio-Visual Contexts: A Study of YouTube Movie Reviews. IEEE Intelligent Systems, 28(3), 46–53.
[2] Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2017). Sentiment Analysis in User-Generated Videos Considering Contextual Factors. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 873–883.
[3] Yoon, S., Byun, S., & Jung, K. (2018). Recognizing Speech Emotion Using Both Audio and Text Modalities. Proceedings of the IEEE Spoken Language Technology Workshop (SLT), 112–118.
[4] Akhtar, M. S., Ekbal, A., & Bhattacharyya, P. (2019). Multi-task Learning Approaches for Emotion Recognition and Sentiment Analysis Using Multiple Modalities. Neurocomputing, 398, 247–260.

AUTHOR'S PROFILES

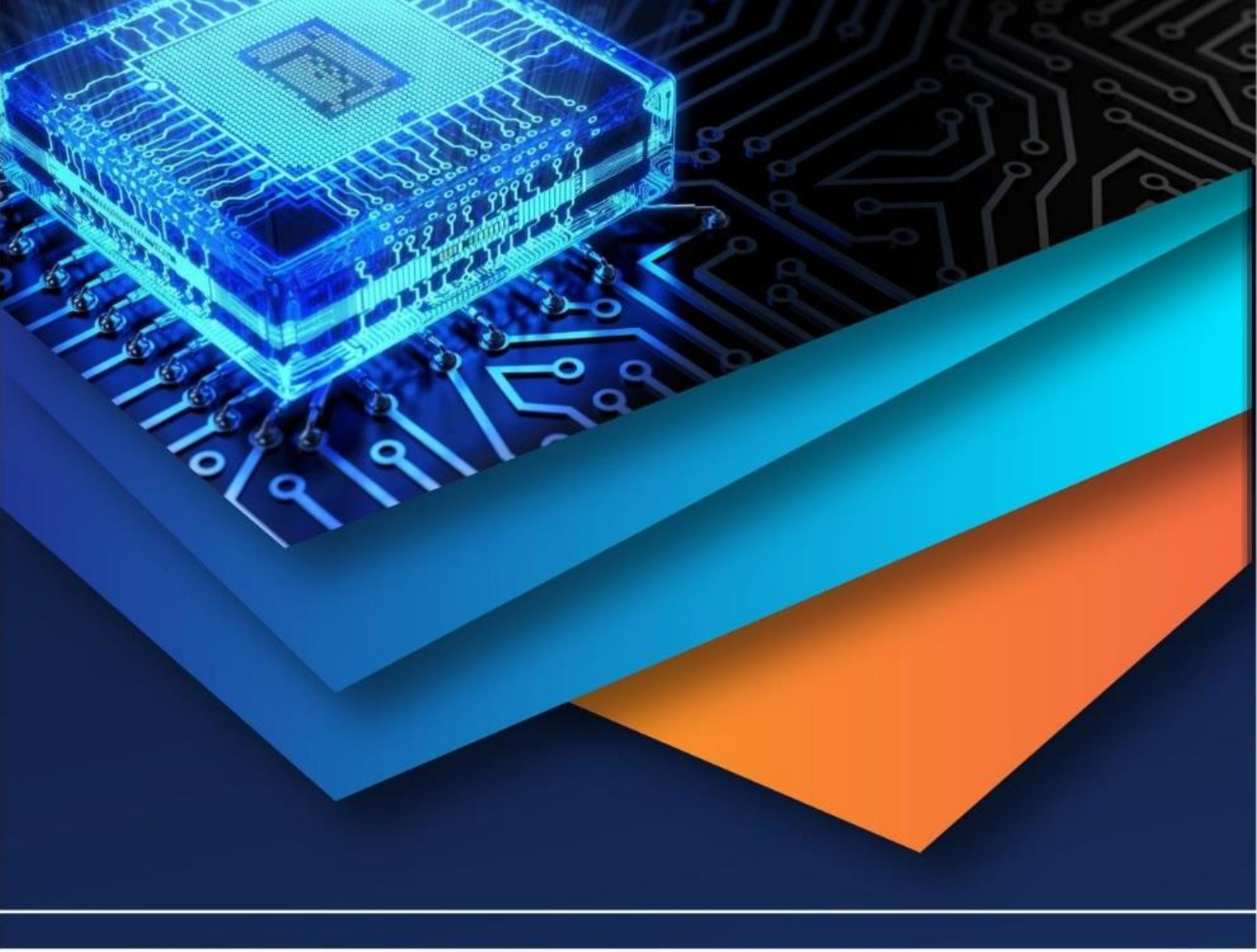
Mrs. G. Venkateswari working as Assistant Professor & HoD in Department of CSE, BWEC, Bapatla.

Y.Srinija B. Tech with specialization of Computer Science and Engineering in Bapatla Women's Engineering College, Bapatla.

P. Sasikala B. Tech with speciliazation of Computer Science and Engineering in Bapatla Women's Engineering College, Bapatla.

T. Hima Deepika B. Tech with specialization of Computer Science and Engineering in Bapatla Women's Engineering College, Bapatla.

P. Sirisha B. Tech with specialization of Computer Science and Engineering in Bapatla Women's Engineering College, Bapatla.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY